

1. (Five Times Amended) A system for determining experimentally a plurality of three-dimensional atomic structures, each of which is associated with a corresponding protein, comprising:

a database of sequence information for a first plurality of proteins, and structural information and functional information for a second plurality of proteins;

at least one bioinformatics tool adapted to use the sequence information, structural information and functional information stored in the database to cluster the first plurality of proteins into a plurality of families, in which, for each family, members of the family have homologous sequences;

protein synthesis means for synthesizing for each family determined by the at least one bioinformatics tool a plurality of target proteins which are members of the family, using information stored in the database corresponding to the target proteins, the protein synthesis means having screening means for screening products of the synthesis to choose selected synthesized products, which are effective as the target proteins, for processing;

protein processing means for preparing, purifying and characterizing each of the selected synthesized products screened through the screening means;

crystallization means for crystallizing the processed synthesized product processed by the protein processing means, to produce a plurality of specimen crystals of the target protein, and testing the plurality of specimen crystals for predetermined diffraction characteristics to determine the specimen crystals which are suitable for diffraction measurement;

X-ray crystallography means for performing high-throughput crystallography on the specimen crystals determined by the crystallization means to be suitable for diffraction measurement,

the X-ray crystallography means having diffraction measuring means for measuring for diffraction data the suitable specimen crystals of the target protein, analyzing means for analyzing the diffraction data, means for building an atomic model of the target protein according to an analysis of the diffraction data by the analyzing means, and means for refining the model of the target protein against the diffraction data and storing the refined model in the database;

structure extraction means having means for analyzing the refined model of the target protein using sequence information corresponding to other family members which is stored in the database and structural information corresponding to other proteins which is stored in the database, and means for analyzing the refined model for functional motifs and for surface characteristics; and

a homology model building tool adapted to use the refined model of the target protein retrieved from the database to develop a homology model of one or more predicted protein structures,

wherein the database is updated using the at least one bioinformatics tool and the developed homology model to link the refined model of the target protein to other databases which store information concerning biological pathways and functional annotation.

7. (Five Times Amended) A process for determining experimentally a plurality of three-dimensional atomic structures, each of which is associated with a corresponding protein, comprising the steps of:

(a) systematically organizing sequence information for a first plurality of proteins, and structural information and functional information for a second plurality of proteins into a

database;

(b) clustering the plurality of proteins into a plurality of families, in which, for each family, members of the family have homologous sequences, using at least one bioinformatics tool and the sequence information, structural information and functional information stored in the database;

(c) synthesizing for each family determined in step (b) a plurality of target proteins which are members of the family, using information stored in the database corresponding to the plurality of target proteins, and screening products of the synthesis to choose selected synthesized products, which are effective as the target proteins, for processing;

(d) preparing, purifying and characterizing each synthesized product that is chosen in step (c);

(e) crystallizing the processed synthesized product prepared, purified and characterized in step (d) to produce a plurality of specimen crystals of the target protein;

(f) testing the plurality of specimen crystals grown in step (e) for predetermined diffraction characteristics to determine the specimen crystals which are suitable for diffraction measurement;

(g) performing high-throughput crystallography, including measuring for diffraction data the specimen crystals determined in step (f) to be suitable for diffraction measurement, building an atomic model of the target protein according to an analysis of the diffraction data, refining the model of the target protein against the diffraction data, and storing the refined model in the database;

(h) analyzing the refined model, stored in the database in step (g), of the target protein using sequence information corresponding to other family members which is stored in the database and structural information corresponding to other

Wayne A. Hendrickson and Barry Honig
Serial No.: 09/235,986
Filed: January 22, 1999
Page B-4

proteins which is stored in the database, and analyzing the refined model of the target protein for functional motifs and for surface characteristics;

(i) developing a homology model of one or more predicted protein structures using computational tools for homology model building and the refined model of the target protein retrieved from the database; and

(j) updating the database by using the at least one bioinformatics tool and the developed homology model to link the refined model of the target protein to other databases which store information concerning biological pathways and functional annotation,

wherein steps (f) through (j) are repeated for each of the other target proteins.